

Une nouvelle perspective et de nouveaux défis pour les enquêtes internationales: vers un *testing* assisté par ordinateur

Romain Martin

Université du Luxembourg

Jean-Guy Blais

Université de Montréal

MOTS CLÉS: Enquêtes comparatives internationales, *testing* assisté par ordinateur, éduométrie

Le présent article analyse les changements qui ont été introduits dans les enquêtes comparatives internationales (plus spécifiquement celles de l'OCDE) par suite de l'introduction du testing assisté par ordinateur. On passera en revue les différents instruments de mesure administrés par ordinateur qui ont été utilisés dans les différents cycles de ces enquêtes et on discutera si elles ont réellement impliqué une valeur ajoutée, notamment par le fait de fournir des données sur des compétences qui sont inaccessibles à travers des instruments classiques sous format papier-crayon. On analysera également les relations spécifiques que les mesures assistées par ordinateur entretiennent avec des variables d'arrière-fond tel le genre. Finalement, l'article mettra en évidence les défis et les risques qui sont associés avec une introduction de plus en plus massive d'instruments de mesure assistés par ordinateur, notamment en ce qui concerne deux aspects majeurs: la nécessité de garder une continuité dans les compétences mesurées en vue de mettre en évidence des évolutions, et le manque de cadres théoriques adéquats pour la mise en œuvre efficace des mesures assistées par ordinateur.

KEY WORDS: Large-scale comparative studies, computer-assisted testing, edumetrics

The present article analyzes the changes that have been introduced in comparative international large-scale studies (specifically those organized by the OECD) as a consequence of the introduction of computer-assisted testing. We will review the various measurement instruments that have been administered by computer in the different cycles of these large-scale studies and we will discuss whether their introduction implied a real added value, notably by providing information on

competencies that are inaccessible through classical paper and pencil tests. We will also analyze the specific relationship that the computer-assisted measures show with background variables, such as gender. Finally, the article will point out the challenges and risks associated with a more and more extensive introduction of computer-assisted measurement instruments, especially regarding two major aspects: the necessity to keep a certain stability in the measured competencies in view of reporting trends, and the lack of adequate theoretical frameworks for the efficient use of the computer-assisted measures.

PALAVRAS-CHAVE: Inquéritos comparativos internacionais, *testing* assistido por computador, edumetria

O presente artigo analisa as mudanças que foram introduzidas nos inquéritos comparativos internacionais (mais especificamente os da OCDE) na sequência da introdução do testing assistido por computador. Passar-se-á em revista os diferentes instrumentos de medida administrados por computador que foram utilizados nos diferentes ciclos destes inquéritos e discutir-se-á se eles implicaram realmente um valor acrescentado, nomeadamente pelo facto de fornecerem dados sobre as competências que são inacessíveis através dos instrumentos clássicos em formato papel e lápis. Analisar-se-á igualmente as relações específicas que as medições assistidas por computador estabelecem com as variáveis de background, como é o caso do género. Finalmente, o artigo evidenciará os desafios e os riscos que estão associados a uma introdução cada vez mais massiva de instrumentos de medida assistidos por computador, designadamente no que respeita a dois aspetos maiores : a necessidade de garantir uma continuidade nas competências medidas com vista a evidenciar as evoluções e a falta de quadros teóricos adequados para a operacionalização eficaz das medições assistidas por computador.

Note des auteurs – Toute correspondance peut être adressée comme suit : Romain Martin, Université du Luxembourg, Faculté des lettres, des sciences humaines, des arts et des sciences de l'éducation, route de Diekirch, L-7220 Walferdange, téléphone : (+352) 46 66 44 9369, télécopieur : (+352) 46 66 44 9556, ou Jean-Guy Blais, professeur titulaire, Département d'administration et fondements de l'éducation, Faculté des sciences de l'éducation, Université de Montréal, C.P. 6128 Succursale Centre-ville, Montréal (Québec) H3C 3J7, ou par courriel aux adresses suivantes : [romain.martin@uni.lu] ou [jean-guy.blais@umontreal.ca].

Introduction

De façon générale, les enquêtes internationales en éducation ont deux objectifs: améliorer la compréhension des systèmes d'éducation et aider à mieux comprendre les causes des résultats et des différences observés entre les systèmes. L'année 2010 célébrait le 50^e anniversaire de la première enquête internationale en éducation menée sous les auspices de l'*International Association for the Assessment of Educational Achievement* (IEA). Cette enquête pilote visait les connaissances des élèves de 13 ans de douze pays pour les mathématiques, la compréhension en lecture, la géographie et la science. Entre 1960 et aujourd'hui, l'IEA a mené régulièrement des enquêtes du même type, les dernières en date étant celles de 2011 sur les mathématiques et la science et sur la lecture (TIMSS et PIRLS). Depuis l'an 2000, l'Organisation de coopération et de développement économiques (OCDE) mène aussi des enquêtes internationales en éducation, qui contribuent significativement à la réflexion sur les caractéristiques des systèmes éducatifs des pays participants. Les enquêtes de l'OCDE, et notamment les enquêtes PISA, sont ainsi devenues des enquêtes comparatives internationales incontournables qui présentent les taux de participation les plus élevés (avec plus de 60 pays participant actuellement à l'enquête PISA 2012) et qui ont l'impact le plus important dans les médias et l'opinion publique.

Avec la première enquête PISA en 2000, l'OCDE a lancé un type d'enquête comparative internationale qui repose sur le concept de littératie, un concept depuis lors utilisé afin de définir les compétences dans des domaines de contenu variés qui sont à la base des instruments de récolte des données mis en œuvre. L'OCDE s'est ainsi explicitement éloignée des descriptions curriculaires ainsi qu'on l'avait vu précédemment dans les enquêtes internationales organisées notamment par l'IEA. La définition des compétences basée sur un dénominateur curriculaire commun identifié dans les pays participant a été abandonnée en faveur d'une définition basée sur la capacité du sujet à mobiliser des processus cognitifs divers qui lui permettent une participation active à une vie sociétale vue avant tout comme le fonctionnement d'une société postindustrielle basée sur une économie de la connaissance.

L'introduction d'un concept de littératie intimement lié aux évolutions des compétences nécessaires pour une participation active à une vie sociétale qui se trouve sous l'emprise de changements technologiques continus et

importants a en même temps généré la nécessité d'adapter les instruments de récolte des données d'une manière continue à ces évolutions sociétales. Cette nécessité d'adaptation des instruments semble constituer un défi d'autant plus grand que les enquêtes de l'OCDE ont été lancées avec la volonté affichée d'accompagner l'évolution des systèmes d'éducation tout en gardant une comparabilité dans le temps des données récoltées. Mais comme l'utilisation du concept de littératie force les enquêtes internationales de l'OCDE à suivre les évolutions sociétales, il n'est pas étonnant de constater qu'elles évoluent de plus en plus pour tenir compte de l'apport des technologies de l'information et de la communication qui va de pair avec la mise en œuvre de compétences clés pour la participation à une société moderne.

Plusieurs éléments sont à considérer lorsqu'on veut aborder la démarche de la mise au point et le déroulement des enquêtes internationales. Il y a d'abord tout ce qui concerne la logistique de l'opération et, donc, l'organisation des séances, la sélection des participants, le développement des tâches à réaliser, le support de ces tâches (papier-crayon ou écran-clavier, par exemple), la correction, la mise en forme des rapports et la diffusion des résultats. Comme l'objectif général des enquêtes est de poser un diagnostic ou de permettre des comparaisons entre les systèmes dans le temps (*trends analysis*), il y a ensuite toute la question de la précision des résultats, donc du contrôle des sources d'erreur comme celles reliées à l'échantillonnage, aux instruments de mesure et aux techniques d'appariement (*equating methods*) des résultats pour différentes cohortes de participants. Finalement, il ne faut pas oublier des éléments politiques collatéraux presque inéluctables comme l'utilisation des résultats à des fins autres que celles pour lesquelles les enquêtes ont été conçues, entraînant des pressions indues sur les systèmes lorsque cela se produit.

À notre avis, l'introduction dans les enquêtes internationales d'épreuves dont la passation se fera en mode écran-clavier permettra une meilleure organisation logistique, l'introduction de tâches plus variées et plus authentiques, un contrôle continu des sources d'erreur de mesure et un ancrage certain avec les évolutions sociétales à venir. Le tout en gardant une cohérence au moins partielle avec les épreuves papier-crayon qui ont été à la base des premières enquêtes internationales. La technologie ne sera cependant pas une panacée et constituera même un des défis majeurs pour l'avenir de ces enquêtes internationales, défi que nous nous proposons d'analyser plus en détail dans le présent article. Vu l'importance actuelle des enquêtes de l'OCDE et du

concept de littératie sur lequel elles sont basées, nous allons, dans notre analyse, prendre aussi comme point d'appui les évolutions actuelles et futures qui se dessinent pour les grandes enquêtes organisées par l'OCDE.

La mise en place progressive des évaluations assistées par ordinateur dans les enquêtes internationales

Le remplacement des épreuves papier-crayon par des passations se déroulant avec un clavier et un écran semble être une révolution qui a été annoncée pour le domaine de l'évaluation en éducation depuis les années 1980 (Bunderson, Inouye, & Olsen, 1989), mais qui n'a jusqu'à aujourd'hui pas véritablement eu lieu. En effet, malgré une ubiquité de l'outil informatique dans tous les domaines de la vie privée ou professionnelle, l'évaluation des connaissances et des compétences est encore avant tout basée sur des outils papier-crayon ou sur la présentation informatisée d'outils qui correspondent à des transpositions directes d'outils existant également sous format papier-crayon (par exemple des items QCM). Néanmoins, la valeur ajoutée potentielle des outils édumétriques assistés par ordinateur est restée un objet de recherche important depuis la réalisation de la première enquête PISA (par exemple, voir Blais, 2009; Blais & Gilles, 2011). En plus, il est évident que les nouvelles compétences qui sont requises par l'évolution sociétale sont quasiment toutes connotées à l'utilisation d'outils technologiques qui font de plus en plus partie intégrante des techniques culturelles que les membres d'une société moderne doivent s'approprier; elles se prêtent ainsi de moins en moins à des formats de tâches en présentation papier-crayon. Il est ainsi logique de constater que dans les vagues successives des enquêtes PISA qui ont suivi l'enquête initiale de 2000, on a de plus en plus essayé d'introduire la passation d'épreuves assistées par ordinateur. Cette évolution a pris son envol définitif avec la troisième vague du PISA, à savoir celle de 2006. Les extensions successives qu'il y a eu ensuite en 2009 et 2012, ainsi que les plans qui existent actuellement pour l'enquête PISA 2015, constituent une bonne illustration pour les évolutions futures des enquêtes internationales et nous allons les présenter plus en détail dans la suite.

L'option CBAS de l'enquête du PISA 2006

L'option CBAS (*computer-based assessment of science*) de l'enquête PISA 2006 a été la première tentative de l'OCDE pour introduire la passation d'une épreuve assistée par ordinateur dans une enquête comparative internationale à large échelle. L'introduction de cette option avait également pour but de préparer le terrain pour une extension future de la passation informatisée, en donnant l'occasion de réaliser pour la première fois des expériences dans le contexte d'une étude comparative à large échelle. L'option CBAS, pour laquelle le rapport final vient d'être publié récemment (Programme for International Student Assessment, 2010b), a d'emblée montré les problèmes d'ordre technique qui sont associés à l'utilisation de ce genre de dispositif dans un contexte de passation à large échelle. Alors que la mise à l'essai sur le terrain a été effectuée en 2005 auprès de treize pays, seulement trois pays ont finalement participé à l'option CBAS principale en 2006 (Danemark, Islande et Corée). Ainsi, afin de fournir un environnement standardisé au maximum, l'épreuve de l'option CBAS a été délivrée sur des ordinateurs portables spécialement équipés et préparés pour la présentation de cette épreuve. Si cette stratégie assure une standardisation optimale de l'environnement informatique, elle implique aussi que le matériel informatique a dû, en partie, être acheté spécifiquement pour cette enquête et qu'il a dû être acheminé dans les écoles pour les séances de passation. Ce transport de matériel signifie un effort logistique considérable et difficilement conciliable avec les contraintes d'une enquête à large échelle (obtenir des données sur un grand nombre de sujets dans une période de temps réduite) où on espère une participation du plus grand nombre de pays ou constituantes éducatives (des provinces, des régions ou des villes). Ces problèmes logistiques constituent ainsi une explication majeure en ce qui concerne le faible taux de participation final à l'étude CBAS.

L'option CBAS était basée sur le même cadre de référence que celui qui a été adopté pour l'épreuve papier-crayon de sciences de PISA 2006 (pour laquelle les sciences étaient le domaine dominant). Cette enquête ne visait donc pas une compétence pour laquelle la passation n'est pas envisageable sans recourir à l'ordinateur et il ne s'agit pas non plus d'une compétence qui est directement liée à l'utilisation d'un dispositif informatisé. Pour cette première étape d'utilisation de l'ordinateur dans une étude internationale à large échelle, on a plutôt choisi de rester dans le cadre de référence de compétences qui pouvaient aussi être évaluées par les instruments classiques

sous forme papier-crayon et d'explorer le potentiel d'extension des tâches à travers les possibilités multimédia offertes par l'ordinateur, ainsi que d'étudier les conséquences en termes de motivation qui sont induites par ce support de présentation nouveau et potentiellement stimulant.

En ce qui concerne l'utilisation des possibilités supplémentaires offertes par l'ordinateur pour étendre l'évaluation des compétences visées par le cadre de référence des sciences, c'est surtout l'intégration d'éléments interactifs dans les stimuli qui a été l'innovation majeure introduite dans l'option CBAS. Le degré d'interactivité qui a été ainsi introduit dans les tâches délivrées par ordinateur était variable, et elles étaient finalement classées en fonction du degré d'interactivité en trois types: tâches à faible, à moyenne et à haute interactivité. Une tâche à faible interactivité peut être illustrée par un item pour lequel le stimulus prend la forme d'une séquence vidéo ou d'une animation qui peut être visualisée par le participant et qui fournit une information contextuelle par rapport aux questions qui lui sont posées. Une tâche à haute interactivité serait par contre illustrée par un item pour lequel le sujet doit interagir avec une simulation afin de trouver des réponses aux questions (p. ex. trouver à travers l'interaction avec une simulation une combinaison optimale entre divers degrés d'acidité du sol et différentes températures ambiantes pour maximiser la croissance d'une plante). Les items de l'option CBAS ont également permis de varier la charge de lecture requise pour les participants, notamment à travers la possibilité offerte par l'ordinateur d'utiliser des vidéos/animations/images à la fois pour l'affichage des stimuli et pour la présentation des options de réponse. Ainsi, les items de l'épreuve de l'option CBAS ont également pu être classés en fonction de leur charge de lecture selon qu'elle était faible, moyenne ou élevée.

Puisque l'étude CBAS utilisait le même cadre de référence que celui qui avait été mis au point pour l'épreuve papier-crayon des compétences en sciences, une première question très intéressante était la stabilité des résultats obtenus avec le support ordinateur par rapport à ceux obtenus avec le support papier-crayon. Cette stabilité s'est effectivement avérée comme étant assez élevée, puisque les corrélations entre les scores de l'option CBAS et les scores papier-crayon en sciences variaient entre 0,78 (pour l'Islande) et 0,89 (pour le Danemark). On a également pu noter que les scores de l'option CBAS étaient un peu moins corrélés avec les scores de lecture (corrélations variant entre 0,68 et 0,77) que ce n'était le cas pour les scores papier-crayon en sciences (corrélations variant entre 0,79 et 0,86). L'utilisation de l'ordinateur pourrait

donc avoir l'effet espéré de pouvoir diminuer la charge en lecture des tâches en sciences et ainsi conférer une meilleure validité à l'évaluation d'une compétence spécifique en sciences.

Les corrélations en général très élevées entre les différentes compétences qui sont visées par les enquêtes PISA constituent d'ailleurs un problème assez important pour la validité des mesures que l'on voudrait réaliser. Le présent exemple illustre en effet que les corrélations entre les scores papier-crayon en sciences et les scores papier-crayon en lecture sont aussi élevées que celles entre les scores papier-crayon en sciences et les scores à l'épreuve de l'option CBAS. Si on prenait donc les scores CBAS en tant que référence pour évaluer le niveau de compétences en sciences et les scores papier-crayon en lecture comme référence pour évaluer le niveau de compétences en lecture, on ne pourrait pas – sur la base des corrélations obtenues – décider si l'épreuve papier-crayon en sciences vise des compétences en sciences ou des compétences en lecture, puisque les valeurs des corrélations trouvées avec les deux épreuves sont comparables. Sur la base de ces constats, on peut donc espérer que les épreuves assistées par ordinateur pourraient au moins fournir une réponse partielle au défi de conférer une meilleure validité discriminante aux résultats qui sont obtenus avec les épreuves du PISA à travers les possibilités nouvelles qui sont ouvertes par l'ordinateur pour réduire la charge de lecture des items qui ne visent pas directement les compétences en lecture. En même temps, l'introduction de premiers éléments interactifs d'investigation scientifique dans le contexte de l'option CBAS a également contribué à augmenter la validité écologique de cette épreuve.

Un autre résultat très intéressant de l'option CBAS était le fait que dans les trois pays ayant participé à l'enquête, les différences de genre en sciences ont évolué d'une manière consistante en faveur des sujets masculins. On a ainsi pu constater qu'une différence d'environ un quart d'écart-type en faveur des sujets masculins s'est ajoutée pour l'option CBAS en comparaison des différences trouvées pour les scores papier-crayon. Ces résultats soulignent donc que les éléments interactifs qui ont été introduits à travers une présentation par ordinateur, ainsi que la diminution de la charge de lecture, ont peut-être une influence différente sur la performance des participants en fonction de leur genre. Puisqu'il est bien connu, notamment à travers les résultats des enquêtes PISA, que les filles ont de meilleures performances en lecture que les garçons, la diminution de la charge de lecture dans l'épreuve informatisée peut être vue comme une hypothèse non négligeable dans l'explication de ce

résultat. Une analyse plus détaillée de cet aspect dans le contexte de l'option CBAS montre qu'une charge de lecture plus importante semble effectivement désavantager certains garçons pour lesquels l'avantage que leur donne les items CBAS joue de manière plus importante lorsque les items ont une charge de lecture qui est faible. Néanmoins, la charge de lecture n'est pas à même d'expliquer à elle seule l'ensemble des différences de genre qui ont été constatées dans les résultats à l'épreuve de l'option CBAS.

On peut également émettre l'hypothèse que la plus grande familiarité que les garçons ont avec l'ordinateur et le plus grand intérêt qu'ils montrent envers cet outil, combinés avec une préférence pour des formats interactifs de présentation de données, contribuent à ces différences de performance en fonction du genre (Martin & Binkley, 2009). En ce qui concerne la relation entre la familiarité avec l'ordinateur et la performance à l'épreuve de l'option CBAS, les analyses correspondantes montrent que le lien entre ces deux variables est complexe. Il s'avère notamment que pour les participants masculins, une plus grande familiarité avec l'outil informatique semble d'une manière consistante aller de pair avec une meilleure performance à l'épreuve, alors que ce résultat ne peut pas être vérifié d'une manière consistante pour les participants féminins. Pour ce qui est de l'interactivité, les résultats à l'épreuve de l'option CBAS n'ont pas pu montrer une influence du degré d'interactivité des tâches sur l'ampleur des différences de genre. Ainsi, une conclusion définitive sur cet aspect sur la base des résultats à l'épreuve CBAS serait certainement prématurée, car même si les tâches ont été classées selon des degrés d'interactivité allant de « faible » à « élevé », l'interaction qui est requise même pour les items des tâches à interaction dite « élevée » reste néanmoins largement en-dessous des possibilités d'interactions dynamiques à l'intérieur d'environnements virtuels complexes qui peuvent en principe être offerts via un ordinateur.

En ce qui concerne l'hypothèse d'une motivation accrue qui est générée par la présentation de l'épreuve sur un support informatisé, cette hypothèse s'est vérifiée pour l'option CBAS. On a pu constater, en effet, que le niveau d'effort investi indiqué par les élèves à la fin de l'épreuve (sur la base d'un questionnaire très court) est en moyenne plus élevé que pour la modalité de passation papier-crayon. On a pu constater, en plus, que les élèves disent avoir investi presque autant d'effort dans la passation de l'épreuve CBAS qu'ils auraient investi dans une épreuve d'évaluation proposée par l'enseignant, alors que ceci n'est pas le cas pour les épreuves papier-crayon.

En conclusion, on peut donc dire que l'enquête avec l'option CBAS a illustré qu'une présentation avec la modalité écran-clavier présente un certain intérêt, mais également un certain nombre de défis. D'un côté, on a effectivement pu montrer qu'une présentation par ordinateur peut contribuer à aborder un éventail plus large de facettes liées à une compétence qui est décrite dans un cadre de référence déterminé tout en augmentant la validité des résultats à travers une diminution de la charge de lecture. En plus, la présentation du test sur support informatisé en ayant recours à des formats de présentation plus variés et plus interactifs semble amener un accroissement général de la motivation des élèves. D'un autre côté, cette nouvelle forme de présentation va de pair avec une modification des différences de genre trouvées précédemment, l'administration des épreuves sur support informatisé semblant jouer en faveur des garçons. Cette évolution peut être saluée si elle correspond véritablement au dévoilement d'une différence de genre bien réelle qui était uniquement cachée par une prépondérance d'éléments n'étant pas directement liés à la compétence visée (par exemple la charge de lecture). Cela dit, le risque est peut-être aussi de remplacer un biais existant par un autre biais, si le format et les contenus de la présentation sur support informatisé sont de nature à correspondre plus fortement aux intérêts et aux motivations des garçons et qu'ils empêchent tout simplement les filles de recourir à leurs modes préférés de traitement de l'information. L'option CBAS doit être vue comme un premier pas en direction d'une exploitation des possibilités offertes par l'ordinateur en vue d'une extension des observations qui peuvent être réalisées dans les enquêtes à large échelle. En effet, l'épreuve de l'option CBAS est malgré tout restée très près des formats classiques et bien connus des instruments papier-crayon et s'est contentée avant tout d'introduire un certain nombre d'éléments d'interactivité dans les nouvelles tâches proposées. Les enquêtes qui vont suivre devraient aller plus loin dans l'exploitation des possibilités de produire une valeur ajoutée rendue possible à travers l'utilisation de l'ordinateur. Néanmoins, l'option CBAS a l'avantage que les résultats issus de cette étude comparative à large échelle (si on peut la considérer comme telle à la suite de la participation de seulement trois pays) ont déjà été publiés, alors que l'ensemble des autres études présentées dans la suite du texte sont encore en cours de traitement de données ou bien en cours d'élaboration, à l'exception de l'option ERA du PISA 2009.

L'option ERA de l'enquête PISA 2009

L'option ERA (*electronic reading assessment*) de l'enquête PISA 2009 se différencie de l'option CBAS avant tout par le fait qu'elle vise une compétence « nouvelle » qui n'est pas couverte par le cadre de référence défini pour la compétence en lecture sous format papier-crayon. Cette compétence nouvelle s'oriente autour des changements dans les habitudes de lecture induits par l'omniprésence de l'outil informatique pour la quête et l'échange d'information, notamment par suite de l'utilisation de plus en plus massive d'Internet et des nouveaux outils technologiques qui y sont associés (tablettes, téléphones intelligents, etc.). Avec l'introduction de l'option ERA, l'OCDE reconnaît donc que les changements sociétaux accompagnant l'introduction des nouvelles technologies de l'information et de la communication (TIC) ont eu un impact sur les techniques culturelles fondamentales telles que la lecture et que cet impact se traduit par la mise en œuvre de processus et de stratégies différents selon le support sur lequel ces techniques culturelles doivent être appliquées. Il est ainsi reconnu que la lecture de textes électroniques exige des préalables et des traitements cognitifs au moins partiellement différents de ceux qui sont mis en œuvre pour la lecture de textes classiques sous format papier-crayon. Or, comme la lecture de textes électroniques occupe une place de plus en plus importante dans les prises d'information quotidiennes, il semble nécessaire de tenir compte de cette évolution également dans les démarches d'évaluation des compétences en lecture (Programme for International Student Assessment, 2010a). Cette exigence de cohérence est d'ailleurs bien soulignée par bon nombre de chercheurs travaillant dans le domaine de l'évaluation des compétences en lecture (Coiro, 2009).

La principale différence entre la lecture de textes sous format papier et la lecture de textes électroniques réside dans le fait que les textes sous format papier présentent un cadre de lecture clairement défini, avec des limites visibles qui guident le lecteur d'une manière séquentielle à travers les différents éléments qui sont constitutifs pour le texte. Les textes papier fournissent donc, à travers leur mode de présentation, un guide pour la mise en œuvre des processus cognitifs nécessaires à l'extraction et au traitement de l'information, alors que ceci est beaucoup moins le cas pour plusieurs des formats actuels de textes électroniques. Une partie de ceux-ci se caractérisent notamment par le fait qu'ils ne présentent pas de limites clairement identifiées, mais plutôt des éléments de navigation dynamiques qui font partie intégrante de ce type de texte et qui deviennent alors également une partie intégrante des stratégies

de lecture qui s'y appliquent. De cette manière, les textes électroniques peuvent impliquer des stratégies de lecture nettement plus discontinues, pour lesquelles les stratégies de navigation et les stratégies de choix des éléments de lecture deviennent un élément important dans la mise au point d'une approche efficace d'extraction et de traitement de l'information en vue de trouver des réponses à des questions données. En plus, les textes électroniques peuvent être enrichis par un ensemble d'éléments multimédia qui est nettement plus grand et plus diversifié que celui qui est envisageable pour des textes sous format papier.

L'objectif de l'option ERA de l'enquête PISA 2009 est donc l'élaboration d'un instrument pour l'évaluation des compétences de lecture de textes électroniques qui présente les caractéristiques qui viennent d'être décrites ci-dessus et qui simule les environnements de lecture de textes électroniques qui peuvent être retrouvés sur Internet et qui sont utilisés par les jeunes (Programme for International Student Assessment, 2010a).

Il n'est donc pas étonnant de voir que les tâches rendues publiques par l'OCDE pour illustrer la lecture de textes électroniques correspondent essentiellement à des simulations de situations de lecture que des jeunes seraient susceptibles de rencontrer sur Internet. À part les formats de lecture typiques pour Internet qui sont caractérisés par des liens de navigation fréquents, on y retrouve également les formats de publication et d'interaction électroniques qui sont au centre des technologies Web 2.0 et qui font de plus en plus partie des expériences de lecture quotidiennes des jeunes : courrier électronique, forums, wikis, blogs, twits, échanges, clavardage (*chat*), etc. Les exemples qui ont été publiés par l'OCDE montrent clairement que les expériences de lecture qui en résultent diffèrent considérablement des expériences de lecture qui caractérisent le format de lecture classique sous format papier. Il semble également évident que les tâches de lecture électronique impliquent un certain nombre de connaissances spécifiques par rapport au support utilisé et touchent donc également, dans une certaine mesure, la familiarité du sujet avec l'outil informatique et les environnements spécifiques au Web 2.0.

Le rapport qui présente les résultats de l'option ERA (OECD, 2011b) semble confirmer le bien-fondé d'une mesure à part pour la lecture de textes électroniques par rapport à la seule évaluation de la compétence de lecture de textes imprimés qui était pratiquée jusqu'ici. En ce qui concerne la validité discriminante, on constate que la corrélation entre les deux types de compétence en lecture (textes électroniques et textes imprimés) est de 0,83. De ce fait,

la corrélation que la lecture de textes imprimés entretient avec la lecture de textes électroniques est bien substantielle, mais pas plus élevée que celle avec la compétence en mathématiques (0,83 également) et même moins importante que celle avec la compétence en sciences (0,88). De son côté, la lecture de textes électroniques entretient avec les mathématiques et les sciences des corrélations qui sont moins importantes que celles trouvées pour la lecture de textes imprimés (0,76 pour la corrélation avec les mathématiques et 0,79 pour la corrélation avec les sciences). La lecture de textes électroniques constitue donc une compétence qui se laisse différencier de celle impliquée dans la lecture de textes imprimés et de celle qui est mobilisée dans le contexte des mathématiques et des sciences.

Un autre résultat très intéressant concerne les différences de genre. Les différences de performance en lecture en fonction du genre étaient, dans toutes les enquêtes du PISA et dans presque tous les pays (ou constituantes éducatives) participants, très clairement en faveur des filles. Or, même si des différences en faveur des filles persistent pour la lecture de textes électroniques, ces différences sont néanmoins nettement moins importantes que celles trouvées pour la lecture de textes imprimés. De même, si on contrôle pour le niveau de compétence en lecture de textes imprimés, les garçons montrent dans la lecture de textes électroniques des compétences partiellement plus élevées que celles des filles. À compétences de lecture de textes imprimés égales, les processus cognitifs impliqués dans la lecture de textes électroniques semblent donc être mieux maîtrisés par les garçons.

Un comportement très important pour la détermination de la performance en lecture de textes électroniques est le comportement de navigation, notamment le nombre de pages pertinentes visitées en vue de donner une réponse à une question spécifique. Il s'est avéré que ce nombre de pages pertinentes visitées entretient une relation forte, mais non linéaire, avec la performance en lecture de textes électroniques. La relation trouvée s'est en effet avérée quadratique, ce qui signifie que la croissance de performance obtenue par les premières pages pertinentes visitées est plus importante que celle qui est obtenue si on augmente encore davantage le nombre de pages pertinentes visitées. En valeur absolue, il s'avère à nouveau que les filles montrent un comportement d'exploration plus efficace que celui des garçons, mais si on contrôle quant au niveau de performance en lecture de textes imprimés, cet avantage en faveur des filles se transforme au moins partiellement dans un avantage en faveur des garçons. Ceci souligne encore une fois que les garçons

semblent avoir une performance moins bonne en ce qui concerne les processus cognitifs directement liés à la lecture, mais qu'ils semblent être relativement plus performants en ce qui concerne les processus spécifiques directement liés à la lecture de textes électroniques.

Ce résultat renvoie également à la question de l'impact de la familiarité avec l'ordinateur sur la performance en lecture de textes électroniques. Le résultat le plus important que l'option ERA produit dans ce contexte doit certainement être vu dans le fait qu'on peut montrer l'existence d'une relation entre ces deux variables, mais que cette relation se révèle encore être non linéaire. On constate ainsi que la fréquence d'utilisation de l'ordinateur à la maison montre une relation curvilinéaire avec la performance en lecture de textes électroniques. Cette relation va donc dans le sens d'une meilleure performance des élèves montrant un taux d'utilisation moyen de l'ordinateur à la maison, alors qu'une faible fréquence d'utilisation tout comme une très forte fréquence d'utilisation semblent être associées avec une performance relativement plus faible en lecture de textes électroniques. Ajoutons encore qu'il y a entre filles et garçons une différence qui abonde dans le sens que les filles qui utilisent l'ordinateur très fréquemment ne dépassent pas la performance de leurs condisciples qui n'utilisent l'ordinateur que rarement. Pour les garçons montrant une fréquence d'utilisation très élevée de l'ordinateur à la maison, on constate de même qu'ils sont moins performants que les garçons montrant une fréquence d'utilisation moyenne, mais ils restent néanmoins plus performants que les garçons qui n'utilisent l'ordinateur que rarement à la maison. Ce résultat indique donc que la lecture efficace de textes électroniques est en partie dépendante d'un entraînement reçu sur le médium de présentation qui a avant tout lieu dans un contexte extrascolaire. Il indique également que les mécanismes qui impliquent ces effets d'entraînement sont plus complexes qu'un simple apprentissage «sur le tas» qui serait uniquement dépendant de la quantité de temps passé sur l'ordinateur. Le fait que les utilisateurs les plus assidus sont au contraire confrontés à une baisse de leur performance renvoie à la nécessité de mieux comprendre les effets de l'utilisation de l'ordinateur sur les processus cognitifs reliés à la lecture afin de pouvoir promouvoir d'une manière plus explicite et plus efficace les processus cognitifs liés à la lecture de textes électroniques dans un contexte scolaire.

Les options de la passation assistée par ordinateur de l'enquête PISA 2012

Pour l'enquête PISA 2012, l'OCDE a encore une fois élargi les domaines pour lesquels des épreuves seront offertes sous une modalité de passation par ordinateur. La dominante sera cette fois-ci l'évaluation des compétences en mathématiques et, pour la première fois, des tâches de ce domaine seront présentées sur support informatique. Néanmoins, si on regarde les ambitions vis-à-vis des tâches présentées par ordinateur, on constate qu'il ne s'agit pas de décrire le niveau d'une compétence qui serait considérée comme qualitativement différente par rapport à celle qui est abordée avec la modalité papier-crayon, une ambition qui était bien présente lors de l'introduction de la compétence pour lire des textes électroniques versus la compétence en lecture de textes classiques (PISA 2012 Mathematics Expert Group, 2010). L'option de présenter les tâches de mathématiques sur support informatisé s'efforce plutôt d'utiliser les possibilités multimédia offertes par l'ordinateur afin d'enrichir des construits qui sont également visés par la modalité papier-crayon. La tâche à accomplir est censée devenir plus motivante pour les élèves, mais par ailleurs on voit déjà dans la définition des ambitions que la valeur ajoutée qui est attendue de cette nouvelle approche reste très incertaine. Cette situation est différente pour une autre compétence qui est visée à l'aide de l'enquête du PISA 2012, à savoir la compétence en résolution de problèmes. Cette compétence avait déjà été évaluée lors du PISA 2003, et le cadre de référence qui a été adopté en 2012 se distingue nettement de celui qui avait été adopté en 2003, notamment par suite de l'introduction de la passation assistée par ordinateur (PISA 2012 Problem Solving Expert Group, 2010). En effet, pour la nature des situations problèmes qui sont visées, on a pu inclure dans le cadre conceptuel de l'enquête du PISA 2012 la résolution de problèmes interactifs, ce qui donne accès à une classe de problèmes qui est décrite dans la documentation sous la dénomination de « problèmes complexes » (Funke, 2010). Ces derniers se caractérisent par un haut degré d'interactivité avec un environnement simulé qui est souvent structuré d'une manière non univoque laissant la place à des objectifs multiples et concurrentiels qui nécessitent de la part du sujet une analyse de l'environnement, une mise au point d'un modèle mental adéquat et une mise en œuvre de stratégies de résolution qui tiennent au mieux compte des différents objectifs que la tâche implique. Les capacités offertes par l'ordinateur pour l'enregistrement de l'ensemble des données comportementales du sujet ouvrent également des possibilités pour dépasser

une simple analyse des résultats en termes de réponses correctes/fausses et d'accéder à une véritable analyse des processus et des stratégies de résolution de problèmes qui ont été mis en œuvre (Greiff & Funke, 2010). L'évaluation de la compétence de résolution de problèmes vise donc encore une fois une mise en œuvre d'une passation assistée par ordinateur avec l'objectif très clair d'étendre conceptuellement le construit visé à travers la nouvelle méthode de passation par rapport à ce qui était visé à travers les instruments papier-crayon. Il s'agit donc ici également d'une tentative d'accéder, à travers des tâches qui peuvent seulement être délivrées par un ordinateur, à l'évaluation de processus cognitifs qui restent inaccessibles aux instruments papier-crayon.

Les plans de l'OCDE pour l'enquête du PISA 2015

Pour l'enquête PISA 2015, la même logique appliquée pour la résolution de problèmes sera appliquée au domaine de l'évaluation des compétences en sciences. Ainsi, l'OCDE demande clairement au consortium qui sera responsable des développements des instruments de l'enquête PISA 2015 d'étendre le cadre de référence de l'évaluation des compétences en sciences afin de tenir compte des possibilités supplémentaires qui sont offertes par l'ordinateur. Par analogie à ce qui a été réalisé pour le cadre de référence relatif à la résolution de problèmes, on cherche donc explicitement à utiliser les possibilités d'interaction et d'affichage enrichi offertes par l'ordinateur pour accroître l'étendue des processus cognitifs reliés à la littératie en sciences qui peuvent être abordés (OECD, 2010, 2011a). On peut donc s'attendre à trouver pour l'enquête PISA 2015 des tâches qui vont inclure par exemple des simulations de phénomènes scientifiques ou des simulations d'environnements d'investigation scientifique. Ces environnements simulés vont être utilisés afin d'obtenir des données sur des processus cognitifs qui sont reliés aux divers aspects du processus d'investigation impliqués dans la réalisation d'une recherche en sciences dans des domaines variés. L'objectif de l'OCDE est donc très clairement d'utiliser la passation assistée par ordinateur afin de dépasser une mesure centrée prioritairement sur les contenus en sciences en faveur d'une mesure tenant plus amplement compte du degré de maîtrise d'une véritable démarche d'investigation scientifique. Remarquons encore que l'enquête PISA 2015 va offrir un large éventail d'épreuves assistées par ordinateur, puisque les mathématiques, la lecture et la résolution de problèmes sont également au programme de cette enquête. Il est intéressant de noter que pour le domaine de la résolution de problèmes, l'OCDE montre encore une fois sa volonté

d'innover par rapport aux compétences évaluées en introduisant la résolution de problèmes collaboratifs. Comme il est très difficile de concevoir comment on pourrait évaluer pour ce domaine innovant les processus cognitifs spécifiques qui y sont associés avec des instruments en format papier-crayon, on se trouve à nouveau confronté à l'obligation d'une passation d'épreuves assistée par ordinateur. Pour l'enquête PISA 2015, l'OCDE prévoit également pour la première fois l'introduction d'un déroulement adaptatif dans les épreuves qui seront proposées, ce qui impliquerait donc de choisir au moins une partie des items sur la base d'une première estimation du niveau de compétence de l'élève. Ainsi on pourrait arriver à une meilleure mise en correspondance de la difficulté des items avec le niveau de compétence de l'élève. Ces éléments adaptatifs sont en effet intéressants en vue de diminuer la variance d'erreur dans les mesures réalisées, mais leur introduction ne se fera certainement pas sans soulever un certain nombre de problèmes méthodologiques (Martin, 2003).

L'étude PIAAC

L'enquête PIAAC (*Programme for the International Assessment of Adults Competencies*) est la première étude comparative internationale de l'OCDE qui est quasiment intégralement basée sur la passation d'épreuves assistées par ordinateur (avec l'exception d'une option de passation papier-crayon de certaines tâches pour les personnes qui n'auraient pas les compétences minimales requises pour la manipulation de base d'un ordinateur). Cette enquête a pour objectif de mesurer les compétences de base dans la population active et s'adresse à des personnes âgées de 16 à 65 ans. Les évolutions constatées pour l'enquête PIAAC en ce qui concerne l'introduction de nouvelles tâches d'évaluation administrées par ordinateur sont en partie analogues à ce qui a été élaboré pour les enquêtes PISA et qui ont été décrites ci-dessus. Ainsi une différence majeure par rapport aux enquêtes précédentes en lecture et visant la même population (enquêtes IALS et ALL), consiste dans le fait qu'on a pu introduire des tâches qui concernent la lecture de textes électroniques (de manière analogue à ce qui a été réalisé pour l'option ERA du PISA) (PIAAC Literacy Expert Group, 2009). Tout comme pour les enquêtes PISA, l'implémentation sur plateforme informatique des épreuves pour les compétences numériques n'a pas vraiment donné lieu, dans l'enquête PIAAC, à l'introduction de construits nouveaux qui seraient uniquement accessibles à travers la passation assistée par ordinateur. On se contente plutôt d'exploiter les

avantages génériques de ce type de passation, telles que les possibilités d’affichage enrichi et de traitement automatique des données (notamment la codification automatique des réponses en provenance du sujet) (PIAAC Numeracy Expert Group, 2009). Mais en principe, les données qui sont obtenues dans ce domaine auraient aussi pu l’être sous format papier-crayon. Il en est tout autrement pour le troisième grand domaine qui est visé par l’enquête PIAAC, à savoir celui de la résolution de problèmes, qui s’appelle d’ailleurs «résolution de problèmes dans un environnement technologiquement enrichi» (PIAAC Expert Group in Problem Solving in Technology-Rich Environments, 2009). Comme la dénomination du domaine l’indique déjà, l’utilisation de l’ordinateur fait ici partie intégrante de la définition du construit de la résolution de problèmes. Ce qu’on envisage pour ce domaine de compétence, c’est de présenter aux sujets des situations problèmes à l’intérieur d’un environnement qui simule un poste de travail classique basé sur l’utilisation de l’ordinateur, c’est-à-dire un environnement qui présente un ordinateur avec accès à Internet et qui présente les logiciels de bureautique classique (courrier électronique, tableur, traitement de textes). On demande alors au sujet de réaliser des tâches spécifiques qui nécessitent l’utilisation soit d’un seul outil, soit d’une combinaison des outils informatiques qui sont présents dans cet environnement de travail. Le choix de ce construit repose sur le fait que le poste de travail qui est présenté s’impose naturellement comme un environnement standard pour les environnements de travail utilisant l’ordinateur et qu’une évaluation des compétences, en ce qui concerne l’utilisation efficace d’un tel environnement, paraît particulièrement importante si on s’intéresse à la population active. On doit néanmoins aussi noter un certain nombre de désavantages qui sont liés à la définition de ce construit :

- 1) dans le cas où les tâches sont surtout basées sur l’utilisation de ressources qui proviennent d’Internet, le construit visé par cette mesure est difficile à distinguer du construit relatif à la lecture de textes électroniques ;
- 2) le construit de résolution de problèmes qui est défini dans le contexte de l’enquête PIAAC est nettement différent de celui qui a été adopté pour l’enquête PISA, ce qui amène des problèmes de cohérence entre les différentes études et ce qui restreint la valeur ajoutée qu’on pourrait attendre d’une mise en relation des données en provenance d’études différentes, mais mesurant des construits analogues ;

- 3) la complétion des tâches implique très fortement la familiarité du sujet avec l'outil informatique et les résultats impliquent donc, à côté des processus cognitifs de résolution de problèmes, également des processus cognitifs reliés à la maîtrise des outils informatiques.

Il reste donc une incertitude sur le fait que ce qui est évalué représente effectivement avant tout une compétence de résolution de problèmes en environnement technologiquement enrichi ou s'il ne s'agit pas plutôt d'une évaluation de la littératie par rapport aux technologies de l'information et de la communication.

L'enquête PIAAC a également introduit des éléments de passation adaptative pour le niveau de compétences du participant, mais pas au niveau des tâches. Il s'agit d'une procédure à branchement basée sur une évaluation initiale des compétences du sujet. Un autre élément notable dans cette enquête en lien avec l'utilisation des TIC est le fait que les questionnaires généraux qui font partie de cette enquête sont administrés sous forme de CAPI (*computer-assisted personal interview*). L'ordinateur est donc utilisé afin d'augmenter l'efficacité de la prise de données et afin d'optimiser la séquence de passation des différents instruments de collecte de données qui sont constitutifs de l'enquête.

Quelques défis techniques pour la validité des comparaisons dans les enquêtes internationales : les apports possibles de la passation par ordinateur

Comme cela a été mentionné au début de ce texte, les enquêtes internationales ont comme objectifs d'améliorer la compréhension des systèmes d'éducation et d'aider à mieux comprendre les causes des résultats et des différences observés entre les systèmes. Ce double souci d'informer et de comparer les pays ou constituantes éducatives qui participent aux enquêtes est louable, certes, mais n'est pas sans soulever des problèmes techniques notamment pour ce qui est de la précision des données récoltées, précision qui est tributaire d'un ensemble d'éléments qui jalonnent les étapes de la récolte et de l'analyse des données (voir à ce sujet le texte de Wu, 2009).

Au départ, la modélisation des scores doit s'appuyer sur des approches qui permettent d'estimer l'erreur de mesure pour chaque participant. Cette modélisation facilite les analyses qui visent à estimer les gains ou pertes observés d'une cohorte à l'autre ou encore permet de suivre les mêmes

participants si l'enquête est longitudinale. De plus, les épreuves plus longues résultent généralement en des erreurs de mesure plus petites pour les scores individuels, mais il y a une limite au nombre de tâches qu'on peut intégrer dans ces épreuves sans étirer indûment les temps de passation et ainsi avoir une influence sur la motivation des participants (qui sont volontaires). Il est fort possible que l'utilisation de l'ordinateur permette la passation d'épreuves avec un plus grand nombre de tâches, mais ce qui contribuera le plus à augmenter la précision de la mesure individuelle (et aussi celle des cohortes) est certainement la possibilité d'introduire des épreuves où les tâches présentées aux participants seront adaptées à leurs connaissances ou à leurs compétences. Des tâches plus faciles pourraient ainsi être soumises à ceux et celles qui ont le plus de difficulté et des tâches plus difficiles à ceux et celles qui ont moins de difficultés selon le domaine visé. Cette adaptabilité des tâches contribue à augmenter la précision de la mesure individuelle pour les enquêtes longitudinales et la précision de la mesure collective pour les enquêtes transversales.

Une deuxième source d'imprécision est celle de l'échantillonnage des individus participants. De façon générale, les responsables des enquêtes internationales essaient, dans la mesure du possible, d'imposer des plans d'échantillonnage communs à tous les pays ou constituantes éducatives qui acceptent de participer. Cependant, il n'est pas rare, pour des raisons logistiques, que certains pays ne puissent satisfaire aux demandes des responsables de cette facette des enquêtes. On le constate en effet dans les rapports produits, plusieurs résultats sont marqués d'astérisques indiquant des précautions à prendre lors des comparaisons en vertu de différences marquées dans les plans d'échantillonnage et donc aussi dans les erreurs d'échantillonnage. Le nombre d'individus sélectionnés joue évidemment un rôle important, mais c'est surtout la façon de les sélectionner qui est primordiale. Ainsi, et ce pour des raisons de convenance, l'échantillonnage se fait en deux étapes dans la majorité des enquêtes internationales: on choisit d'abord un certain nombre d'écoles au hasard et ensuite une ou deux classes dans cette école. Ce plan d'échantillonnage introduit une corrélation interne (*intraclass correlation*) qui augmente l'erreur d'échantillonnage et contribue à diminuer la précision des estimations par rapport à ce qui serait obtenu avec un échantillon aléatoire simple. Lorsque la participation est sur une base volontaire, la passation à l'aide de l'ordinateur pourrait être une source de motivation supplémentaire à participer ou encore permettre de rejoindre plus facilement les individus. Cependant, malgré les apports possibles de la technologie pour la passation des épreuves, le plan d'échantillonnage (et donc l'erreur qui y est associée) demeure la res-

pensabilité des pays et constituantes éducatives participants qui ne peuvent pas toujours garantir le respect des exigences en la matière et qui ne peuvent non plus garantir la volonté de participer des écoles et des individus sélectionnés.

Une troisième source d'imprécision, et peut-être la plus importante, concerne la volonté de suivre le progrès des systèmes d'éducation sur de longues périodes de temps. Deux obstacles importants se dressent devant cet objectif. D'abord, et en prenant l'exemple des enquêtes PISA, si des changements notables sont introduits dans les tâches à réaliser, il faut se demander si le construit visé, et donc l'essence de la mesure, demeure stable à travers ces changements lorsque ceux-ci deviennent de plus en plus importants au fur et à mesure que l'on s'éloigne du repère temporel initial retenu pour la comparaison et que l'on utilise de plus en plus de tâches à interaction élevée (par exemple le construit « sciences » de 2006 sera-t-il toujours le même en 2021?). Ensuite, lorsqu'on désire faire ces comparaisons longitudinales ou transversales, il faut établir un design d'appariement (*equating design*). Ce design d'appariement fait appel à des items d'ancrage communs et aussi à des modèles de mesure qui peuvent être assez sophistiqués techniquement. Jusqu'à maintenant, comme l'intégration d'une erreur reliée à l'appariement dans l'étude des tendances temporelles n'a pas été considérée de façon systématique, l'influence du non-respect des conditions d'application de ces modèles sur les résultats de l'appariement a aussi été peu étudiée. Et ce, malgré le fait que l'erreur reliée à l'appariement contribue fortement, par exemple, aux différences d'estimation des moyennes de cohortes dans les comparaisons transversales (Michaelides & Haertel, 2004). Ainsi, Monseur et Bereznier (2007) ont montré, à partir d'une étude sur le fonctionnement différentiel d'items (FDI) de l'épreuve de l'enquête PISA 2003, que le retrait d'un seul item problématique à cet égard de l'ensemble des items d'ancrage pour l'appariement faisait varier de dix points la moyenne d'un des pays participants, une différence assez élevée pour modifier le rang du pays concerné. La contribution d'une passation par ordinateur pourrait ici être plus indirecte. D'une part, cette passation permet la production d'un très grand nombre d'épreuves différentes avec des items d'ancrage communs, augmentant ainsi les points possibles pour les comparaisons tout en introduisant une meilleure prise en compte des problèmes reliés au modèle retenu pour l'appariement des scores. D'autre part, comme les résultats d'une telle passation sont obtenus immédiatement, des procédures spécifiques pour l'étude des biais potentiels peuvent aussi être introduites plus rapidement.

Conclusion

Pour conclure, nous pouvons retenir que les grandes enquêtes internationales subissent actuellement un processus de transformation majeure, notamment dû à l'introduction de la passation d'épreuves informatisées. Au vu des évolutions qu'il y a eu lors des derniers cycles de ces grandes enquêtes internationales et de la place de plus en plus importante qu'on y a allouée aux passations d'épreuves sur support informatisé, il paraît dorénavant clair que la passation du « tout papier-crayon » au « tout informatisé » est une volonté affichée de l'OCDE et que cette transition va se réaliser dans les années à venir. Or, même si cette transition fait sens au vu des évolutions sociétales, et notamment des évolutions dans la vie professionnelle auxquelles on a pu assister au cours des dernières décennies, elle est néanmoins très loin de ne pas comporter de considérables risques et périls.

Ainsi on peut douter du fait que tous les pays qui participent actuellement aux enquêtes internationales vont pouvoir fournir, avec la même facilité et au même horizon temporel, l'infrastructure nécessaire pour des passations à large échelle d'épreuves sur ordinateur. Une solution intermédiaire pourra alors résider dans le fait d'offrir, pendant une période de transition, des épreuves disponibles sous deux formats différents (papier-crayon et ordinateur), ce qui comportera néanmoins deux désavantages majeurs. D'abord, on met alors en cause l'équivalence des épreuves qui sont utilisées dans des pays différents, car même si les travaux dans ce domaine ont montré une assez bonne équivalence des épreuves fournies sur des supports différents (*cf.* Mead & Drasgow, 1993), on peut néanmoins douter de l'absence totale d'effets de présentation et mettre ainsi en péril un des aspects essentiels des enquêtes internationales, à savoir la comparabilité entre pays. Ensuite, si les épreuves doivent se prêter à une passation équivalente sous format papier-crayon et sur ordinateur, on se restreint du côté ordinateur à des types d'items qui ne peuvent pas tirer avantage des possibilités d'affichage enrichie et des possibilités d'interaction offertes par ordinateur, ce qui veut dire qu'on perd une raison majeure de l'introduction même des épreuves assistées par ordinateur.

Un autre point essentiel qui va toucher un des objectifs majeurs des enquêtes internationales sera celui de l'ancrage des mesures réalisées dans le temps afin de pouvoir suivre l'évolution des systèmes éducatifs. Ce suivi des évolutions dans le temps sera possible à la seule condition que les construits mesurés restent stables dans le temps. Or, on est en droit d'exprimer des doutes par rapport à la stabilité des construits qui étaient initialement mesurés

exclusivement par des épreuves papier-crayon si on introduit de plus en plus d'épreuves délivrées par ordinateur qui vont impliquer des éléments interactifs auxquels seul l'ordinateur va pouvoir donner accès. Est-ce que cette introduction d'une interaction plus complexe va engendrer un changement de la nature des construits qu'on va mesurer et, si oui, comment sera-t-on encore à même de parler d'une évolution sur le plan des pays, puisque les mesures réalisées lors des premiers cycles des enquêtes internationales ne seront plus vraiment comparables à celles qui sont réalisées dans les cycles les plus récents ?

Un dernier point critique qui doit être souligné concerne plus directement les nouveaux instruments de mesure interactifs assistés par ordinateur qui vont être introduits au cours des prochains cycles des enquêtes internationales. Force est en fait de constater que justement par rapport à la valeur ajoutée que ces nouveaux instruments sont censés fournir (notamment les comportements interactifs à l'intérieur d'environnements complexes), on manque très souvent de modèles de mesure adéquats qui permettent de pleinement mettre à profit cette valeur ajoutée potentielle qui semble être accessible à travers des données comportementales largement enrichies par rapport au support papier-crayon. L'introduction d'une épreuve de résolution de problèmes collaboratifs pour le PISA 2015 peut être vue comme un bon exemple pour illustrer ce propos. Pour ce domaine, il manque en effet à la fois un cadre théorique bien établi qui puisse décrire et délimiter ce champ et il manque notamment des modèles de mesure qui puissent permettre l'établissement de critères de performance pour des épreuves collaboratives dans un contexte d'évaluation à large échelle et à caractère comparatif. Ceci illustre donc le risque d'introduire d'une manière précoce des mesures qui paraissent réalisables et innovantes d'un point de vue technique, mais en négligeant la réalisabilité d'un point de vue psychométrique et d'un point de vue de la compréhension théorique des processus cognitifs et conatifs impliqués.

En dépit des critiques qui viennent d'être formulées, il ne faut néanmoins pas oublier que les épreuves assistées par ordinateur comportent un potentiel bien réel en vue d'enrichir les grandes enquêtes internationales. Or, pour véritablement pouvoir exploiter ce potentiel, il paraît indispensable de réaliser encore des efforts de recherche importants au cours des années à venir dans le domaine des épreuves assistées par ordinateur. Ces efforts de recherche devront clarifier les questions de comparabilité entre épreuves papier-crayon et épreuves assistées par ordinateur et devront surtout établir la valeur ajoutée

des mesures interactives réalisées sur support informatique par rapport aux épreuves papier-crayon. Ces recherches devront combiner différents domaines scientifiques:

- 1) les sciences de l'information, en vue de réaliser des environnements innovants à l'intérieur desquels la collecte de données va avoir lieu ;
- 2) la psychométrie, qui devra établir des modèles de mesure valides pour la multitude de données comportementales qui seront collectées à l'aide de l'ordinateur ; et
- 3) les sciences cognitives, qui devront fournir une description adéquate des processus cognitifs et motivationnels qui seront impliqués dans les nouvelles épreuves présentées sur ordinateur.

En fait, il nous semble que de voir aboutir ce type de recherches est une condition *sine qua non* pour que les enquêtes internationales puissent à l'avenir tirer pleinement profit de la nouvelle ère des épreuves assistées par ordinateur.

RÉFÉRENCES

- Blais, J.-G. (dir.) (2009). *Évaluation des apprentissages et technologies de l'information et de la communication : Enjeux, applications et modèles de mesure*. Québec : Les Presses de l'Université Laval.
- Blais, J.-G., & Gilles, J.-L. (dir.) (2011). *Évaluation des apprentissages et technologies de l'information et de la communication : Le futur est à notre porte*. Québec : Les Presses de l'Université Laval.
- Bunderson, V. C., Inouye, D. K., & Olsen, J. B. (1989). The four generations of computerized educational measurement. In R. L. Linn (dir.), *Educational measurement: Third edition* (p. 367-407). New York: Macmillan.
- Coiro, J. (2009). Rethinking online reading assessment. *Educational Leadership*, 66(6), 59-63.
- Funke, J. (2010). Complex problem solving: A case for complex cognition? *Cognitive Processing*, 11(2), 133-142 [doi:10.1007/s10339-009-0345-0].
- Greiff, S., & Funke, J. (2010). Systematic investigation of complex problem-solving systems based on minimal complex. *Zeitschrift Fur Padagogik*, 216-227.
- Martin, R. (2003). Le testing adaptatif par ordinateur dans la mesure en éducation: Potentialités et limites. *Psychologie et psychométrie*, 24(2-3), 89-116.
- Martin, R., & Binkley, M. (2009). Gender differences in cognitive tests: A consequence of gender-dependent preferences for specific information presentation formats? In F. Scheuermann & J. Björnsson (dir.), *The transition to computer-based assessment: New approaches to skills assessment and implications for large-scale testing* (p. 75-81). Luxembourg: Office for Official Publications of the European Communities.
- Mead, A. D., & Drasgow, F. (1993). Equivalence of computerized and paper-and-pencil cognitive ability tests: A meta-analysis. *Psychological Bulletin*, 114(3), 449-458.
- Michaelides, M. P., & Haertel, E. H. (2004). *Sampling of common items: an unrecognised source of error in test equating*. Centre for the study of evaluation, CRESST, University of California at Los Angeles.
- Monseur, C., & Berezner, A. (2007). The computation of equating errors in international surveys in education. *Journal of Applied Measurement*, 8(3), 323-335.
- OECD (2010). *Development of the cognitive frameworks for the sixth cycle of the programme for international student assessment (PISA 2015)*. Paris: OECD Call for Tenders.
- OECD (2011a). *Development of the computer platform & development of the context questionnaires and their framework for the sixth cycle of the programme for international student assessment (PISA 2015)*. Paris: OECD Call for Tenders.
- OECD (2011b). *PISA 2009 results: Students on line: Digital technologies and performance (volume VI)*. Paris: OECD Publishing [doi:http://dx.doi.org/10.1787/9789264112995-en].
- PIAAC Expert Group in Problem Solving in Technology-Rich Environments (2009). *OECD Education Working Papers: Vol. 36. PIAAC problem solving in technology-rich environments: A conceptual framework*. Paris: OECD Publishing [doi:10.1787/220262483674].
- PIAAC Literacy Expert Group (2009). *OECD Education Working Papers: Vol. 34. PIAAC literacy: A conceptual framework*. Paris: OECD Publishing [doi:10.1787/220348414075].

- PIAAC Numeracy Expert Group (2009). *OECD Education Working Papers: Vol. 35. PIAAC numeracy: A conceptual framework*. Paris: OECD Publishing [doi:10.1787/220337421165].
- PISA 2012 Mathematics Expert Group (2010). *PISA 2012 mathematics framework*. Paris: OECD Publishing.
- PISA 2012 Problem Solving Expert Group (2010). *PISA 2012 field trial problem solving framework*. Paris: OECD Publishing.
- Programme for International Student Assessment (2010a). *PISA 2009 assessment framework: Key competencies in reading, mathematics and science*. Paris: OECD Publishing.
- Programme for International Student Assessment (2010b). *PISA computer-based assessment of student skills in science*. Paris: OECD Publishing.
- Wu, M. (2009). *Issues in large scale assessment*. Texte présenté à la conférence Pacific Rim Objective Measurement Symposium, juillet, Hong Kong.

Date de réception : 13 octobre 2011

Date de réception de la version finale : 24 octobre 2011

Date d'acceptation : 25 octobre 2011